Research

Elucidating Gene-by-Environment Interactions Associated with Differential Susceptibility to Chemical Exposure

Michele Balik-Meisner,[1]*[†] Lisa Truong,[2]*[‡] Elizabeth H. Scholl,[1]*[†] Jane K. La Du,[2]*[‡] Robert L. Tanguay,[2]*[‡] and David M. Reif[1]*[†]

[1]Bioinformatics Research Center, Center for Human Health and the Environment, Dept. of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA
[2]Sinnhuber Aquatic Research Laboratory, Dept. of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon, USA

BACKGROUND: Modern societies are exposed to vast numbers of potentially hazardous chemicals. Despite demonstrated linkages between chemical exposure and severe health effects, there are limited, often conflicting, data on how adverse health effects of exposure differ across individuals.

OBJECTIVES: We tested the hypothesis that population variability in response to certain chemicals could elucidate a role for gene–environment interactions (GxE) in differential susceptibility.

METHODS: High-throughput screening (HTS) data on thousands of chemicals in genetically heterogeneous zebrafish were leveraged to identify a candidate chemical (Abamectin) with response patterns indicative of population susceptibility differences. We tested the prediction by generating genome-wide sequence data for 276 individual zebrafish displaying susceptible (Affected) vs. resistant (Unaffected) phenotypes following identical chemical exposure.

RESULTS: We found GxE associated with differential susceptibility in the *sox7* promoter region and then confirmed gene expression differences between phenotypic response classes.

CONCLUSIONS: The results for Abamectin in zebrafish demonstrate that GxE associated with naturally occurring, population genetic variation play a significant role in mediating individual response to chemical exposure. https://doi.org/10.1289/EHP2662

Introduction

Little is known about why the effects of environmental agents, including industrial chemicals, manufacturing by-products, metals, pesticides, and herbicides differ between individuals within and across populations (Abdo et al. 2015; National Academies of Sciences, Engineering, and Medicine et al. 2016; Blaser et al. 2013). There is strong evidence that gene–environment interactions (GxE) play an important role in health outcomes, and that these interactions are likely a major source of the heterogeneity in chemical response. It is well established that GxE are an important component of the etiology of complex traits and diseases (Hunter 2005). Pharmacogenomic studies have consistently demonstrated that differential susceptibility to chemical exposure is directly related to genetic variation (Johnson 2003; Motsinger-Reif et al. 2013). Thus, for any genetically diverse population, latent genetic variation may contribute to observed differential susceptibility when challenged by chemical exposure. While it is accepted that such GxE interactions are important, there remain significant challenges—both experimental and statistical—in detecting and characterizing such interactions (Rappaport and Smith 2010; Zeise et al. 2013).

The zebrafish (*Danio rerio*), with over 26,000 protein-coding genes consisting of orthologues for over 70% of human genes, has gained momentum as a model organism in vertebrate genomics (Howe et al. 2013; Lieschke and Currie 2007). Toxicological and pharmacological applications in chemical biology have also seized upon the many benefits of zebrafish, including the short generation time, well-characterized development, and early maturation as clear embryos (Kimmel et al. 1995). Various morphological end points are easy to observe, and effects of multiple chemical exposures on these outcomes have been broadly studied (Asharani et al. 2011; Bai et al. 2010; Truong et al. 2014; Usenko et al. 2007). These advantages have led to an upward trend in high-throughput zebrafish chemical screens, especially toward screens of many chemicals in many fish, primarily in 96-well plates (Rennekamp and Peterson 2015). Thus, there exists potential for large-scale studies of chemical bioactivity that integrate genetic information to probe mechanisms underlying morphologic response to chemical exposures during development (Baer et al. 2014) or even across multiple generations (Knecht et al. 2017; Kovács et al. 2015).

Zebrafish populations differ from many model organisms in that the standard husbandry practices can be designed to maintain diversity (Nasiadka and Clark 2012), meaning that most laboratory populations contain an unknown level of genetic diversity (Brown et al. 2012). While this diversity is attractive in translating to questions of human and ecological health, it raises critical questions of how unmeasured interindividual genetic variation might contribute to susceptibility differences in response to chemical exposure. Uncharacterized genetic diversity can manifest as apparent error effects within and across laboratories (Rennekamp and Peterson 2015).

Comparisons between named strains and interlab populations of zebrafish have shown variability in several phenotypes, providing the rationale that constitutive genetic variation may contribute to the variability in exposure response (Lange et al. 2013). Unfortunately, partitioning this variability among genetic, environmental, and phenotypic factors is hindered by (non)systematic differences in experimentation, statistical analysis, and most importantly, by a lack of available genetic data for the strains evaluated. Despite the small samples (one to two individual fish) or pooled strategies used in studies aiming to characterize genetic diversity, results have shown between 5 and 15 million single-nucleotide polymorphisms (SNPs) segregating in a zebrafish population, with roughly half of the variants showing evidence of population specificity (Butler et al. 2015; LaFave et al. 2014; Obholzer et al. 2012; Patoway et al. 2013).

To address these limitations, we directly interrogated GxE interactions by individually sequencing entire genomes from a large sample of zebrafish using refined experimental and statistical methods for characterizing phenotypic responses to chemical exposure (Garcia et al. 2016; Truong et al. 2016; Zhang et al. 2016, 2017). The analytical methods were developed using years of data from high-throughput studies HTS of diverse chemicals in the Tanguay Laboratory's (Oregon State University, Corvallis, OR) Tropical 5D Zebrafish line (T5D). The T5D line is an outbred population of heretofore unknown genetic heterogeneity that has been used to screen thousands of chemicals for adverse biological responses (Reif et al. 2016; Truong et al. 2014). We leveraged these HTS data to identify a chemical (Abamectin) that elicited biological response patterns indicative of population genetic differences, and then generated genome-wide sequence data to compare individuals displaying differential susceptibility to chemical exposure. To our knowledge, this is the first genome-wide association study (GWAS) using individually sequenced zebrafish drawn from a diverse population. Our approach inverts the paradigm of typical GxE research, where rather than interrogating the same list of usual chemical suspects, we only ask questions of particular compounds that have strong evidentiary support for differential population susceptibility.

Methods

Overview of Experimental Approach

First, we exploited the large-scale design of a high-throughput screening (HTS) system that has tested thousands of chemicals in an outbred zebrafish population of unknown genetic heterogeneity to select Abamectin as a chemical that displayed phenotypic patterns of high variability between individual responses. Second, we characterized the correlation structure across morphological end points in order to describe a specific multivariate Affected vs. Unaffected phenotype in zebrafish exposed to Abamectin. Third, we performed range-finding studies to narrow our estimate of the Abamectin critical concentration to that at which a stable 50:50 Affected:Unaffected proportion was observed. Fourth, samples were exposed to the critical concentration of Abamectin, from which we isolated DNA from Affected samples that displayed our multivariate phenotype and Unaffected samples that did not respond to chemical exposure. Fifth, whole-genome sequences were generated from DNA individually isolated from each Affected and Unaffected sample. These data were then used to identify GxE as genetic variants associated with differential response to chemical exposure.

Experimental Population and Developmental Screening System

Adult T5D (wildtype) zebrafish were housed at Sinnhuber Aquatic Research Laboratory at Oregon State University. All experiments in this manuscript used this population, which originated from a minimum of 25 small group crosses, each group containing three males and three females. Each generation was propagated using equal proportions of offspring from the original group crosses. Adult zebrafish were group spawned to produce embryos for the developmental screening system detailed in Truong et al. (2014). In this system, embryos were dechorionated (Truong et al. 2011) and placed into individual wells of a 96-well plate. Chemical exposures were initiated at 6 hpf and evaluated at 120 hpf (hours post fertilization) for 17 morbidity and mortality end points. The end points evaluated were MORT: mortality at 120 hpf, YSE: yolk sac edema, AXIS: bent body axis, EYE: eye, SNOU: snout, JAW: jaw, OTIC: otic, PE: pericardial edema, BRAI: brain, SOMI: somite, PFIN: pectoral fin, CFIN: caudal fin, CIRC: circulation, PIG: pigmentation, TRUN: trunk length, SWIM: swim bladder, and NC:

notochord distortion. Each end point was recorded as present or absent for each individual embryo. These binary vectors of 17 morphological end points per individual were used for subsequent analysis.

Identifying Chemicals with Evidence of Differential Population Response

Screening data on 1,060 chemicals from Truong et al. (2014) were evaluated to prioritize chemicals for GWAS mapping. Our aim was to identify the chemical with maximal evidence of differential response for optimal GxE power. Chemical choice was based on morphological data from a large-scale, concentration–response study of chemically exposed zebrafish embryos assessed for developmental toxicity end points at 120 hpf. In searching for patterns indicative of population variability over a broad, multipoint concentration series, our prioritization metric highlighted maximal population variability in response across all binary morphological end points to identical environmental exposures. The full concentration–response data for 1,060 chemicals ($n = 32$ samples tested across each of five chemical concentrations) were analyzed. We first produced a sublist of chemicals that had at least one morphological end point (other than mortality) observed at near 50% incidence (32–68%) at a minimum of two concentrations. This measure ensured that there was developmental variability between fish exposed to that chemical (i.e., high proportions of both Affected and Unaffected individuals), that a stable critical concentration could be estimated, and that the response variability held for more than one concentration. Given our goal of maximizing interindividual response variability, the latter principle ensured a less steep concentration–response curve and a higher probability that reproducible patterns of differential population response to exposure would be observed in subsequent trials.

Generating Data at a Critical Exposure Concentration to Elicit Maximal Population Variability

Two rounds of range-finding experiments were performed to estimate a critical concentration of test chemical (Abamectin, CASRN 71751-41-2, Sigma-Aldrich) that induced 50% incidence of effect so that the genetic association study could draw evenly from Affected and Unaffected individuals. For each round, adult zebrafish were group spawned to produce embryos that were exposed to several concentrations of Abamectin, following the developmental screening protocol detailed above. First, 576 individual zebrafish embryos were distributed into six 96-well plates, with 12 individuals per plate exposed to a concentration (0, 0.03, 0.1, 0.3, 1, 3, 5, or 10 μM) using a 20-mM stock of Abamectin digitally dispensed using an HP D300e (Hewlitt Packard HP D300e). Second, a narrower range including 0, 0.5, 0.7, and 0.9 μM was performed using 192 individuals (48 per concentration). From these data, the final critical concentration of 0.6 μM was selected to approximate a 50% effect size while ensuring sufficient numbers of completely clean (i.e., absence of adverse developmental end points) Unaffected samples.

After the concentration was established, individual zebrafish were exposed to the critical concentration of 0.6 μM of Abamectin to identify individuals responding differently to equivalent environmental concentrations. We also included untreated (negative) controls on each plate (Exposed: control ratio at 72:24 per plate) to ensure that we could detect global plate effects and have confirmatory samples to sequence if unexpected genotype distributions had been encountered. The total number of samples exposed was 768 (8 plates with 96 individuals), including 576 Exposed:192 Control. From the exposed embryos and knowledge of end point–end point relationships and developmental cascades (Zhang et al. 2016, 2017), we identified samples of maximum phenotypic clarity (i.e.,

no intermediates) to define a clean multivariate phenotype of individuals Affected by exposure vs. individuals Unaffected by the same Abamectin exposure. For Affected status, an individual had to display all of the following specific end points: altered eye, snout, jaw, and axis development, plus pericardial and yolk sac edema (EYE, SNOU, JAW, AXIS, PE, and YSE). The Unaffected phenotypic status was applied to individual embryos having absence of any morphological defect (i.e., normally developed embryos). A total of 276 (138 Affected, 138 Unaffected) zebrafish exposed at the critical concentration were randomly selected to be sequenced for genome-wide association mapping. While other individuals may have been classifiable into one of our two extreme, Affected/Unaffected classes, we optimized sequencing resources towards fully penetrant phenotypes in order to maximize discriminatory power.

Genotyping by Sequencing

Genomic DNA was extracted from individual larvae (Quick-DNA™ 96-Kit, Cat # D3011; Zymo Research), with a subset selected for sequencing. For each sample, 350 ng of DNA was used for library preparation. Prior to library prep, the quality and quantity was verified using a fluorometric plate reader and bioanalyzer. Samples were sheared to ~320 base pair (bp), and 100 ng of each sample was used in the WaferGen DNA amplification library prep kit (Takara Bio, USA). After the library prep, each sample was quantified to verify similar input for sequencing. The samples were sequenced on an HiSeq3000 (Illumina) with 12 samples per lane (~five times coverage) and 150-bp paired-end sequencing. All library preparation and sequencing was performed at Oregon State University's Center for Genome Research and Biocomputing (http://cgrb.oregonstate.edu/core).

Quality Control and Alignment

FastQC (Babraham Bioinformatics version 0.11.3) tests for sequence quality, Guanine Cytosine (GC) content, sequence length distribution, sequence duplication levels, overrepresented sequences, adaptor content, and kmer content were used to ensure fidelity of results. For each sample (DNA from an individual zebrafish), reads were aligned to the Genome Reference Consortium GRCz10 (Howe et al. 2013) reference genome with Bowtie2 (version 2.1.0; open source) (Langmead and Salzberg 2012), using standard settings. Potential PCR duplicates were then removed using Samtools rmdup (Li et al. 2009).

Variant Calling and Filtering

Variant calls were generated for each individual at every variant site. A variant call was made at any site (across the entire genome, including all chromosomes and mitochondrial DNA, excluding nonchromosomal material or scaffolds not aligned within a chromosome) where there was sufficient evidence of a nonreference base for at least one individual. GATK version 3.5 (McKenna et al. 2010) HaplotypeCaller was used to call genotypes on all samples simultaneously (joint genotyping). This leverages data across samples to assign genotypes for individuals with low coverage at certain bases using a Bayesian likelihood model for genotyping. Reads with a mapping quality (MQ) below 20 were not included, and a minimum phred-scaled confidence threshold of 10 was required. Genotypes are reported for every individual at every variant site for which there remained reads.

The GATK VariantFiltration tool was used to implement the GATK Best Practices (DePristo et al. 2011) hard filtering recommendations (filter SNPs with quality by depth <2; phred-scaled Fisher's exact test p-value >60; root mean score MQ <35; MQ Mann-Whitney rank sum <-12.5; read position Mann-Whitney rank sum <-8; strand odds ratio >3). The MQ threshold of 35 was adjusted from GATK's recommendation of 40. This is due to the GATK workflow's use of a different aligner, which outputs a larger range of MQ scores for each base, averaging 60 for high confidence reads. The maximum MQ outputted from the Bowtie2 aligner is only 42 (for a perfectly aligned read with no mismatches to the reference). A final filtering refinement and file conversion for subsequent analysis was performed using VCFtools [VCFtools version 0.1.14 (Danecek et al. 2011)]. SNPs with mean depth (across samples) below 2 were excluded. This was done to reduce false-positive calls for SNPs that were based on too few reads per individual.

Association Analysis

An allelic association analysis (Fisher's exact tests) was run for each SNP across the genome using PLINK 1.9 (Purcell et al. 2007). For each SNP, the counts of the minor and major alleles in the cases (morphologically Affected individuals) and controls (morphologically Unaffected individuals) were used to calculate the exact hypergeometric probability of observing those four counts under the null hypothesis that allele counts in cases and control do not differ. A p-value below $\alpha = 0.05$ provides sufficient evidence that the allele counts for each group do statistically differ. To adjust for multiple testing, a Bonferroni correction was used, leading to a significant call for any SNP with a p-value $< 2.55 \times 10^{-9}$ (0.05/19,597,672).

Validation

For validation of candidate SNPs identified in our GWAS, gene expression was measured following Abamectin exposure on a new set of individuals that were randomly sampled from the population. The developmental screening protocol was identical to that described earlier. At 120 hpf, 84 individuals (balanced between control, Affected, and Unaffected) were randomly selected for real-time polymerase chain reaction (RT-PCR) gene expression analysis. Embryos from wells of interest (Affected and Unaffected) were independently snap frozen using liquid nitrogen and then copurified for RNA and DNA using ZR-Duet™ MiniPrep (Cat #D7001; Zymo Research). A list of the forward and reverse primers (targeting a SNP just upstream of *sox7*, which had the smallest GWAS p-value) used for the experiment can be found in Table S1 (Integrated DNA Technologies). The mRNA gene expression analysis was performed as described in Chlebowski et al. (2017).

Expression of target gene *sox7* and housekeeping gene *beta actin* were measured as threshold cycle (C_T) value using relative standard curves to optimize how much input was in the reaction. Prior to analysis of RT-PCR results, one control individual was discarded for an abnormally high C_T value ($C_T > 32$; for other samples, $27 < C_T < 30$), and six individuals (two Affected; four Unaffected) were discarded due to insufficient amplification signal. A two-sample t-test was first conducted to ensure that *beta actin* was not differentially expressed in control individuals compared to exposed (Affected and Unaffected) individuals ($p = 0.58$). The statistical analysis of *sox7* expression was then based on the $\log_2$ fold change calculated for each individual using the $\Delta\Delta C_T$ method:

$$\Delta C_T(i) = sox7\ C_T(i) - beta\ actin\ C_T(i)$$

$$\Delta\Delta C_T(i) = \Delta C_T(i) - \sum_{j=1}^{c} \frac{\Delta C_T(j)}{c}$$

$$Fold\ Change(i) = 2^{-\Delta\Delta C_T(i)}$$

Individual i is one of the total number of control, Affected, and Unaffected individuals, and c is the number of control individuals. The fold change for each individual is with respect to the

average control. A two-sample *t*-test was then conducting comparing $\log_2$ (fold change) for Affected vs. Unaffected individuals.

Results

Response Patterns Indicative of Differential Susceptibility

We exploited the HTS data from Truong et al. (2014) to identify chemicals with empirical evidence of population susceptibility differences, as illustrated in Figure 1A. Out of 1,060 chemicals screened in full concentration–response by Truong et al. 2014, there were 19 that met our population variability heuristic (Table S2). From this empirical shortlist, Abamectin had the most robustly variable response (i.e., highest proportion of responders) across the

most concentrations. Given the broad concentration spacing ($\log_{10}$) of the HTS design, certain individuals (Unaffected) tolerated concentrations of Abamectin that were orders of magnitude higher than concentrations causing severe abnormalities in other individuals (Affected).

Rangefinder Experiments to Pinpoint a Critical Concentration

Figure 1B illustrates the progression of rangefinder experiments aimed at identifying the critical concentration for morphological effects induced by Abamectin. Power estimates showed that in the absence of prior knowledge of allele frequencies or population genetic structure, the optimal study design should include
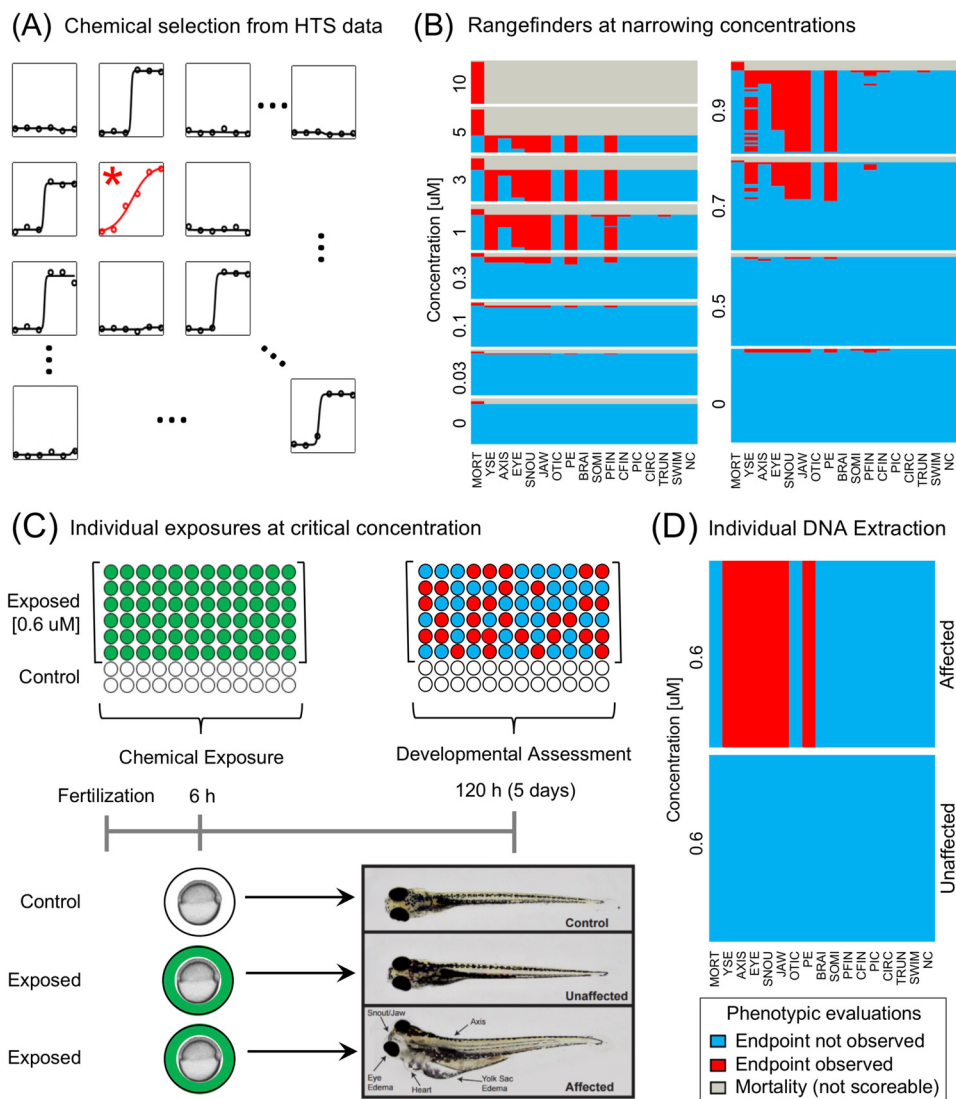
Figure 1. Study Design. (*A*) Chemical selection from HTS data: Example concentration–response curves from 1,060 chemicals interrogated for adverse morphological end points. Each panel represents a test chemical where the proportion of individuals displaying adverse morphological development (vertical axis) is plotted against the tested concentrations (horizontal axis). The curve with the asterisk in the upper left-hand corner of the panel represents a chemical response suggestive of differential population susceptibility, whereas all other curves depict steeper toxic points of departure (i.e., less spread in the range of concentrations eliciting effects across the population) or lack of response. (*B*) Rangefinders: Successive screens to find the critical concentration as that at which approximately 50% incidence is observed. The heatmaps show horizontal blocks (separated by whitespace) of identical concentrations, whose height corresponds to the number of zebrafish tested. Within each concentration block, each row is the vector of observed morphological end points (17 columns; see "Methods" section) for an individual. As per the legend at the lower right, blue represents no end point incidence, red represents incidence of an end point, and grey cross-hatching represents mortality. (*C*) Critical concentration exposure: Example of a single exposure plate, where 72 individuals (in single wells) were exposed to 0.6 μM Abamectin at 6 hpf, plus 24 individuals exposed to vehicle [dimethylsulfoxide (DMSO)] controls. Developmental morphology screening was performed at 120 hpf to identify Affected individuals with the phenotype of altered eye (EYE), snout (SNOU), jaw (JAW), pericardial edema (PE), yolk sac edema (YSE), and axis development (AXIS) vs. Unaffected individuals with no observed defects. (*D*) Individual DNA extraction: Individuals classified as Affected and Unaffected were randomly selected for whole-genome sequencing.

a balanced phenotypic ratio of Affected:Unaffected samples. Therefore, rangefinders aimed to find the concentration eliciting an even ratio of our complete phenotype. The first round narrowed the variable–response concentrations of the original HTS data to a range between 100 ng. The second round tightened the target range between 0.03–10 μM. From these data, 0.6 μM was chosen as the critical concentration as intermediate between the high incidence at 0.7 μM and the low incidence at 0.5 μM.

Identifying Individuals for Genomic Sequencing

Of the 576 fish exposed to Abamectin at the critical concentration, we observed 3% mortality at the 120 hpf evaluation. Of the surviving exposed fish, 155 (28%) displayed the fully penetrant Affected phenotype (see Figure 1C, image of Affected individual with arrows pointing to specific end points), 200 fish (36%) were Unaffected, and the remainder of surviving individuals were scored as having an intermediate phenotype that consisted of a subset of the full Affected end point set. This is evidence of population variability in response to chemical exposure. From the fish at the two phenotypic extremes, 276 individuals were randomly chosen for full-genome sequencing (138 Affected and 138 Unaffected). By first focusing on a clean Affected group of fish that scored exactly the same on the morphologic measures, we reduced potential variability coming from sources other than genetics and thereby increased the power to detect association. In unexposed control individuals, only 1% showed any specific morphological deformity (none of whom showed the specific phenotype of interest), vs. 65% of exposed individuals ($p < 10^{-16}$).

Genetic Polymorphisms Associated with Gene-by-Environment

FastQC output indicated that reads were 151 bp in length. The overall alignment rate was 89% for each sample. GC content for each sample was approximately 37%, which is consistent with the zebrafish genome (Han and Zhao 2008). All individual samples passed QC and were retained. Before base quality control/filtration, there were 44,150,378 variant sites (36,532,474 SNPs) with an average of 4.2 times coverage per site. After applying filtering cutoffs, 19,973,683 SNPs remained. The final VCFtools filter left 19,597,672 SNPs for association analysis.

Association analysis was conducted for each SNP across the genome using Fisher's exact test in PLINK 1.9 (Purcell et al. 2007) to assess allele counts in the Affected vs. Unaffected individuals. To adjust for multiple testing, we subjected our nominal p-value ($\alpha = 0.05$) to a Bonferroni correction of (0.05/19,597,672), yielding a genome-wide significance threshold of $p < 2.55 \times 10^{-9}$. The strict statistical significance criteria highlighted three SNPs that exceeded the genome-wide significance thresholds (Figure 2 and Table S3). The SNPs (ranked in order of ascending p-value) were mapped to genic regions of *sox7*, *erf*, and *cfap74*.

Validation

From the genome-wide data, a G → T variant in the promoter region (569 bp upstream) of *sox7* (the top hit in Figure 2) was significantly associated with severe developmental end points after exposure. This SNP had the smallest p-value and was observed at high prevalence in the population tested, highlighting a promising
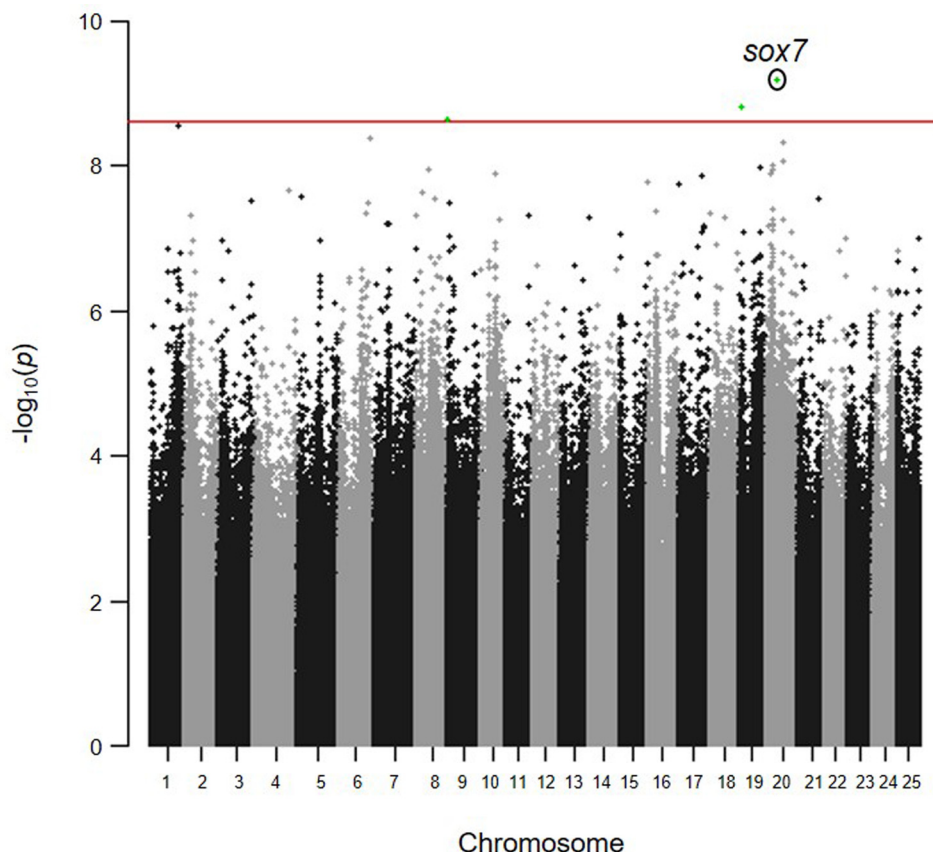
Figure 2. Genome-wide association study (GWAS) results for Abamectin. The Manhattan plot shows the genomic coordinate for each SNP on the horizontal axis (grouped into chromosomes) vs. its association with phenotypic status on the vertical axis (as the negative logarithm of p-value). The horizontal line indicates the Bonferroni-adjusted significance threshold. The dots above this line indicate candidate SNPs (*cfap* on chromosome 8, *erf* on chromosome 19, *sox7* on chromosome 20) for validation as genetic factors associated with differential susceptibility (i.e., Affected vs. Unaffected phenotypes) to Abamectin exposure.
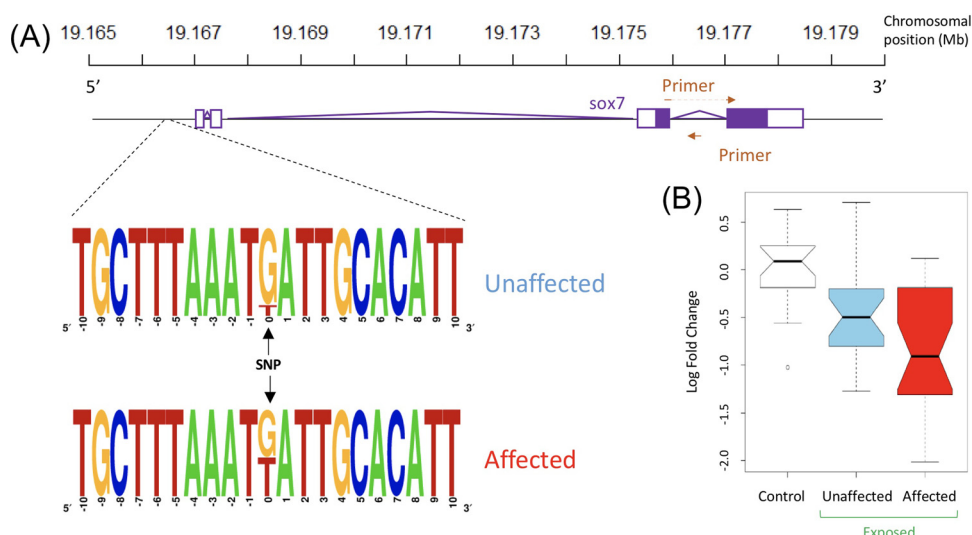
Figure 3. Functional Validation of *sox7*. (*A*) Depiction of the *sox7* transcript, gene expression primer locations, and frequency sequence logos for the region surrounding the significant SNP (20:19,166,444) in Affected and Unaffected individuals from the genome-wide association study (GWAS). Sequence logos are centered at the SNP site, denoted as position 0. The relative letter height corresponds to the frequency of the base at that position. (*B*) Notched boxplots showing the distribution of $\log_2$ (fold change) of *sox7* expression for Control (unexposed), Unaffected (exposed), and Affected (exposed) groups. The boxplots show the median (thick horizontal lines), the upper and lower quartiles (top and bottom of boxes, respectively), 1.5 times the interquartile range (whiskers), and any outlier samples outside the whisker range of the observed expression for each group.

target for functional validation. The SNP region upstream of *sox7* and expression primer design for the validation study are highlighted in Figure 3A and Table S1. The lower portion of Figure 3A shows the T allele frequency differences at this SNP between Affected (45%) and Unaffected (12%) individuals. In the validation study, using a new sample of individuals from the population, *sox7* showed significantly lower expression ($p = 0.02$; Figure 3B) at 120 hpf for Affected ($n = 26$) vs. Unaffected individuals ($n = 24$) after exposure to 0.6 μM Abamectin.

Discussion

Our evidence from Abamectin-exposed zebrafish showed that interindividual (i.e., population) genetic variation contributes to differential response to environmental chemical exposures. To reach this conclusion, we first exploited the large-scale, systematic design of HTS data (Truong et al. 2014) to select a target chemical (Abamectin) whose exposure produced patterns of differential response in the exposed population. Next, we generated genome-wide sequence data for individual zebrafish displaying susceptible vs. resistant phenotypes following identical chemical exposure. Finally, we identified a genetic region near *sox7* associated with this GxE effect and confirmed gene expression differences between susceptibility groups.

This approach addresses a critical need in the face of an expanding chemical exposome (Wild 2005). Select individuals or entire communities may be especially susceptible to adverse health effects from chemical exposure through common consumer products, occupational hazards, environmental emergencies, or geographic location, such as Superfund sites (Tilley et al. 2017; Brette et al. 2014; Judson et al. 2010). Models for diverse populations are needed to explore this interindividual susceptibility (French et al. 2015). The T5D zebrafish line used in our experiments presents such a genetically heterogeneous population, containing levels of interindividual diversity required for studies of differential susceptibility (Balik-Meisner et al. 2018).

In contrast to the pooled samples commonly used for these types of experiments in zebrafish (Butler et al. 2015; Obholzer et al. 2012), we followed individuals (in single-fish wells) from immediately postfertilization throughout the entire environmental exposure course, all phenotypic assessments, and generation of genetic

information. Analysis of individual-level differences in behavioral and morphological responses to chemicals provided solid phenotypic anchors for genetic results. Indeed, our results demonstrated reproducible population variability in a multivariate phenotype that showed consistent concentration–response to chemical treatment across successive rounds of narrowing concentration. The fine-scale quantification of phenotype and exposure environment enabled us to elucidate the role that genetic variation can play in differential susceptibility. This specificity, where we can identify individual-level genetic variation that affects response to individual chemical environments, may bring new precision to personalized toxicity prediction and risk assessment.

Importantly, the authors note that the test compound, Abamectin, is not a genotoxic compound that would have been expected to alter DNA sequence (Oliveira et al. 2016). Abamectin is a standard compound formulation of avermectin B1a and B1b and member of the structurally complex mectin class of compounds. It is used to control insects in agriculture by acting through glutamate-gated chloride channels [Gamma-aminobutyric acid (GABA) receptor] and as an anthelmintic agent to treat common intestinal worms (Campbell 1989). Abamectin has evidence of population variability in response to pharmaceutical applications (Aljedani and Almehmadi 2016; Churcher et al. 2009; Khaldoun-Oularbi et al. 2013; Slimko et al. 2002).

A novel SNP upstream of *sox7* was associated with GxE at a genome-wide significance level. There is strong evidence that this gene, a transcription factor, plays a critical role in development related to our Affected phenotype. Ablation of *Sox7* in mice leads to developmental delays, pericardial edema, and yolk sac defects (Wat et al. 2012). In zebrafish, *sox7* mutants have arterial block and pericardial edema after 72 hpf (Hermkens et al. 2015). Our functional validation experiments showed statistically significant suppression of *sox7* gene expression in Affected individuals vs. those Unaffected following chemical exposure. Although these expression differences were observed in a new, unbiased sample of individuals, there are likely more factors at play in diverting exposed individuals from normal development toward a severely abnormal phenotype. Additional experiments will be needed that probe key time points along the coordinated cascade of vertebrate

developmental events (Zhang et al. 2017). We present the current results as a step toward unraveling this etiology by highlighting promising genetic candidates for detailed study.

Further experimentation must be undertaken before we can completely unravel the causative etiology of genetic associations with differential susceptibility. Indeed, initial deep sequencing efforts indicated that the *sox7* SNP identified by our GWAS is highly correlated with nearby insertion/deletion variation. Thus, as with any genome-wide scan, the markers identified as significant highlight regions that, when considered in tandem with biological knowledge, are candidates for targeted follow-up. For example, the second SNP passing our significance threshold fell within the *erf* gene, for which there is evidence from other species for a plausible connection to our Affected phenotype. An *Erf* mutation in humans and mice is linked to craniosynostosis and related eye, snout, and jaw deformities (Twigg et al. 2013).

In order to provide a clear test of whether our system could detect GxE related to differential population susceptibility, we implemented a straightforward approach of well-characterized genetic analysis tools. By focusing on the reduction of measurement error in phenotypes and having knowledge of an individual's entire exposure history, we augmented our statistical detection power. However, this does not discount the possibility that higher-order interactions between genetic factors (i.e., epistasis) are involved (Motsinger et al. 2007). Future work will include methods aimed at assessing this gene–gene interaction space (Zhao et al. 2015).

Conclusions

Our approach leverages whole-organism, HTS data to introduce a data-driven strategy for identifying chemicals with putative GxE effects. This search for empirical patterns of differential response as an indicator of possible genetic explanations diverges from studies that aim to further characterize the small subset of chemicals with demonstrated GxE effects. Only with strong evidentiary support for differential population susceptibility do we assay the role of genetics in response to particular compounds. Moreover, such an unbiased approach may identify candidate GxE chemicals from a broader search space than traditionally studied, more accurately representing the diversity of real-world chemical space. In conclusion, by linking experimentation with bioinformatical prediction, we can make more informed choices on new experimental directions and avoid unnecessary expenditures of time and money-chasing effects that are unlikely to reflect genetic variation.

Acknowledgments

References

Abdo N, Xia M, Brown CC, Kosyk O, Huang R, Sakamuru S, et al. 2015. Population-based in vitro hazard and concentration–response assessment of chemicals: the 1000 genomes high-throughput screening study. Environ Health Perspect 123(5):458–466, PMID: 25622337, https://doi.org/10.1289/ehp.1408775.

Aljedani DM, Almehmadi RM. 2016. Effects of some insecticides on longevity of the foragers honey bee worker of local honey bee race Apis mellifera jemenatica. Electron Physician 8(1):1843–1849, PMID: 26955457, https://doi.org/10.19082/1843b.

Asharani PV, Lianwu Y, Gong Z, Valiyaveettil S. 2011. Comparison of the toxicity of silver, gold and platinum nanoparticles in developing zebrafish embryos. Nanotoxicology 5(1):43–54, PMID: 21417687, https://doi.org/10.3109/17435390.2010.489207.

Baer CE, Ippolito DL, Hussainzada N, Lewis JA, Jackson D, Stallings JD. 2014. Genome-wide gene expression profiling of acute metal exposures in male zebrafish. Genom Data 2:363–365, PMID: 26484131, https://doi.org/10.1016/j.gdata.2014.10.013.

Bai W, Zhang Z, Tian W, He X, Ma Y, Zhao Y, et al. 2010. Toxicity of zinc oxide nanoparticles to zebrafish embryo: a physicochemical study of toxicity mechanism. J Nanopart Res 12(5):1645–1654, https://doi.org/10.1007/s11051-009-9740-9.

Balik-Meisner M, Truong L, Scholl EH, Tanguay RL, Reif DM. 2018. Population genetic diversity in zebrafish lines. Mamm Genome 29(1-2):90–100. 1, PMID: 29368091, https://doi.org/10.1007/s00335-018-9735-x.

Blaser M, Bork P, Fraser C, Knight R, Wang J. 2013. The microbiome explored: recent insights and future challenges. Nat Rev Microbiol 11(3):213–217, PMID: 23377500, https://doi.org/10.1038/nrmicro2973.

Brette F, Machado B, Cros C, Incardona JP, Scholz NL, Block BA. 2014. Crude oil impairs cardiac excitation-contraction coupling in fish. Science 343(6172):772–776, PMID: 24531969, https://doi.org/10.1126/science.1242747.

Brown KH, Dobrinski KP, Lee AS, Gokcumen O, Mills RE, Shi X, et al. 2012. Extensive genetic diversity and substructuring among zebrafish strains revealed through copy number variant analysis. Proc Natl Acad Sci USA 109(2):529–534, PMID: 22203992, https://doi.org/10.1073/pnas.1112163109.

Butler MG, Iben JR, Marsden KC, Epstein JA, Granato M, Weinstein BM. 2015. SNPfisher: tools for probing genetic variation in laboratory-reared zebrafish. Development 142(8):1542–1552, PMID: 25813542, https://doi.org/10.1242/dev.118786.

Campbell WC, ed. 1989. *Ivermectin and Abamectin*. New York, NY: Springer.

Chlebowski AC, Garcia GR, La Du JK, Bisson WH, Truong L, Massey Simonich SL, et al. 2017. Mechanistic investigations into the developmental toxicity of nitrated and heterocyclic PAHs. Toxicol Sci 157(1):246–259, PMID: 28186253, https://doi.org/10.1093/toxsci/kfx035.

Churcher TS, Pion SD, Osei-Atweneboana MY, Prichard RK, Awadzi K, Boussinesq M, et al. 2009. Identifying sub-optimal responses to ivermectin in the treatment of River Blindness. Proc Natl Acad Sci USA 106(39):16716–16721, PMID: 19805362, https://doi.org/10.1073/pnas.0906176106.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158, PMID: 21653522, https://doi.org/10.1093/bioinformatics/btr330.

DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491–498, PMID: 21478889, https://doi.org/10.1038/ng.806.

French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, et al. 2015. Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. Environ Health Perspect 123(3):237–245, PMID: 25376053, https://doi.org/10.1289/ehp.1408202.

Garcia GR, Noyes PD, Tanguay RL. 2016. Advancements in zebrafish applications for 21st century toxicology. Pharmacol Ther 161:11–21, PMID: 27016469, https://doi.org/10.1016/j.pharmthera.2016.03.009.

Han L, Zhao Z. 2008. Comparative analysis of CpG Islands in four fish genomes. Comp Funct. Genomics 2008:565631, PMID: 18483567, https://doi.org/10.1155/2008/565631.

Hermkens DM, van Impel A, Urasaki A, Bussmann J, Duckers HJ, Schulte-Merker S. 2015. Sox7 controls arterial specification in conjunction with hey2 and efnb2 function. Development 142(9):1695–1704, PMID: 25834021, https://doi.org/10.1242/dev.117275.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496(7446):498–503, PMID: 23594743, https://doi.org/10.1038/nature12111.

Hunter DJ. 2005. Gene–environment interactions in human diseases. Nat Rev Genet 6(4):287–298, PMID: 15803198, https://doi.org/10.1038/nrg1578.

Johnson JA. 2003. Pharmacogenetics: potential for individualized drug therapy through genetics. Trends Genet 19(11):660–666, PMID: 14585618, https://doi.org/10.1016/j.tig.2003.09.008.

Judson RS, Martin MT, Reif DM, Houck KA, Knudsen TB, Rotroff DM, et al. 2010. Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. Environ Sci Technol 44(15):5979–5985, PMID: 20602150, https://doi.org/10.1021/es102150z.

Khaldoun-Oularbi H, Richeval C, Djenas N, Lhermitte M, Humbert L, Baz A. 2013. Effect of sub-acute exposure to abamectin (insecticide) on liver rats (Rattus norvegicus). Ann Toxicol Anal 25(2):63–70, https://doi.org/10.1051/ata/2013039.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203(3):253–310, PMID: 8589427, https://doi.org/10.1002/aja.1002030302.

Knecht AL, Truong L, Marvel SW, Reif DM, Garcia A, Lu C, et al. 2017. Transgenerational inheritance of neurobehavioral and physiological deficits from developmental exposure to benzo[a]pyrene in zebrafish. Toxicol Appl

Pharmacol 329:148–157, PMID: 28583304, https://doi.org/10.1016/j.taap.2017.05.033.

Kovács R, Csenki Z, Bakos K, Urbányi B, Horváth Á, Garaj-Vrhovac V, et al. 2015. Assessment of toxicity and genotoxicity of low doses of 5-fluorouracil in zebrafish (Danio rerio) two-generation study. Water Res 77:201–212, PMID: 25889180, https://doi.org/10.1016/j.watres.2015.03.025.

LaFave MC, Varshney GK, Vemulapalli M, Mullikin JC, Burgess SM. 2014. A defined zebrafish line for high-throughput genetics and genomics: NHGRI-1. Genetics 198(1):167–170, PMID: 25009150, https://doi.org/10.1534/genetics.114.166769.

Lange M, Neuzeret F, Fabreges B, Froc C, Bedu S, Bally-Cuif L, et al. 2013. Inter-individual and inter-strain variations in zebrafish locomotor ontogeny. PLoS One 8(8):e70172, PMID: 23950910, https://doi.org/10.1371/journal.pone.0070172.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359, PMID: 22388286, https://doi.org/10.1038/nmeth.1923.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079, PMID: 19505943, https://doi.org/10.1093/bioinformatics/btp352.

Lieschke GJ, Currie PD. 2007. Animal models of human disease: zebrafish swim into view. Nat Rev Genet 8(5):353–367, PMID: 17440532, https://doi.org/10.1038/nrg2091.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303, PMID: 20644199, https://doi.org/10.1101/gr.107524.110.

Motsinger AA, Ritchie MD, Reif DM. 2007. Novel methods for detecting epistasis in pharmacogenomics studies. Pharmacogenomics 8(9):1229–1241, PMID: 17924838, https://doi.org/10.2217/14622416.8.9.1229.

Motsinger-Reif AA, Jorgenson E, Relling MV, Kroetz DL, Weinshilboum R, Cox NJ, et al. 2013. Genome-wide association studies in pharmacogenomics: successes and lessons. Pharmacogenet Genomics 23(8):383–394, PMID: 20639796, https://doi.org/10.1097/FPC.0b013e32833d7b45.

National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Board on Life Sciences. 2016. *Interindividual Variability: New Ways to Study and Implications for Decision Making: Workshop in Brief*. 30 September–1 October 2015, Washington, DC. Washington, DC: The National Academies Press, 1–13, https://doi.org/10.17226/23413.

Nasiadka A, Clark MD. 2012. Zebrafish breeding in the laboratory environment. ILAR J 53(2):161–168, PMID: 23382347, https://doi.org/10.1093/ilar.53.2.161.

Obholzer N, Swinburne IA, Schwab E, Nechiporuk AV, Nicolson T, Megason SG. 2012. Rapid positional cloning of zebrafish mutations by linkage and homozygosity mapping using whole-genome sequencing. Development 139(22):4280–4290, PMID: 23052906, https://doi.org/10.1242/dev.083931.

Oliveira R, Grisolia CK, Monteiro MS, Soares AM, Domingues I. 2016. Multilevel assessment of ivermectin effects using different zebrafish life stages. Comp Biochem Physiol Part C Toxicol Pharmacol 187:50–61, PMID: 27153811, https://doi.org/10.1016/j.cbpc.2016.04.004.

Patowary A, Purkanti R, Singh M, Chauhan R, Singh AR, Swarnkar M, et al. 2013. A sequence-based variation map of zebrafish. Zebrafish 10(1):15–20, PMID: 23590399, https://doi.org/10.1089/zeb.2012.0848.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575, PMID: 17701901, https://doi.org/10.1086/519795.

Rappaport SM, Smith MT. 2010. Environment and Disease Risks. Science 330(6003):460–461, PMID: 20966241, https://doi.org/10.1126/science.1192603.

Reif DM, Truong L, Mandrell D, Marvel S, Zhang G, Tanguay RL. 2016. High-throughput characterization of chemical-associated embryonic behavioral changes predicts teratogenic outcomes. Arch Toxicol 90(6):1459–1470, PMID: 26126630, https://doi.org/10.1007/s00204-015-1554-1.

Rennekamp AJ, Peterson RT. 2015. 15 years of zebrafish chemical screening. Curr Opin Chem Biol 24:58–70, PMID: 25461724, https://doi.org/10.1016/j.cbpa.2014.10.025.

Slimko EM, McKinney S, Anderson DJ, Davidson N, Lester HA. 2002. Selective electrical silencing of mammalian neurons in vitro by the use of invertebrate ligand-gated chloride channels. J Neurosci 22(17):7373–7379, PMID: 12196558, https://doi.org/10.1523/JNEUROSCI.22-17-07373.2002.

Tilley SK, Reif DM, Fry RC. 2017. Incorporating ToxCast and Tox21 datasets to rank biological activity of chemicals at Superfund sites in North Carolina. Environ Int 101:19–26, PMID: 28153528, https://doi.org/10.1016/j.envint.2016.10.006.

Truong L, Reif DM, St Mary L, Geier MC, Truong HD, Tanguay RL. 2014. Multidimensional in vivo hazard assessment using zebrafish. Toxicol Sci 137(1):212–233, PMID: 24136191, https://doi.org/10.1093/toxsci/kft235.

Truong L, Bugel SM, Chlebowski A, Usenko CY, Simonich MT, Simonich SL, et al. 2016. Optimizing multi-dimensional high throughput screening using zebrafish. Reprod Toxicol 65:139–147, PMID: 27453428, https://doi.org/10.1016/j.reprotox.2016.05.015.

Truong L, Harper SL, Tanguay RL. 2011. Evaluation of embryotoxicity using the zebrafish model. Methods Mol Biol 691:271–279, PMID: 20972759, https://doi.org/10.1007/978-1-60761-849-2_16.

Twigg SR, Vorgia E, McGowan SJ, Peraki I, Fenwick AL, Sharma VP, et al. 2013. Reduced dosage of erf causes complex craniosynostosis in humans and mice and links erk1/2 signaling to regulation of osteogenesis. Nat Genet 45(3):308–313.

Usenko CY, Harper SL, Tanguay RL. 2007. In vivo evaluation of carbon fullerene toxicity using embryonic zebrafish. Carbon N Y 45(9):1891–1898, PMID: 18670586, https://doi.org/10.1016/j.carbon.2007.04.021.

Wat MJ, Beck TF, Hernández-García A, Yu Z, Veenma D, Garcia M, et al. 2012. Mouse model reveals the role of SOX7 in the development of congenital diaphragmatic hernia associated with recurrent deletions of 8p23.1. Hum Mol Genet 21(18):4115–4125, PMID: 22723016, https://doi.org/10.1093/hmg/dds241.

Wild CP. 2005. Complementing the genome with an exposome: the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14(8):1847–1850, PMID: 16103423, https://doi.org/10.1158/1055-9965.EPI-05-0456.

Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ Health Perspect 121(1):23–31, PMID: 23086705, https://doi.org/10.1289/ehp.1205687.

Zhang G, Marvel S, Truong L, Tanguay RL, Reif DM. 2016. Aggregate entropy scoring for quantifying activity across endpoints with irregular correlation structure. Reprod Toxicol 62:92–99, PMID: 27132190, https://doi.org/10.1016/j.reprotox.2016.04.012.

Zhang G, Roell KR, Truong L, Tanguay RL, Reif DM. 2017. A data-driven weighting scheme for multivariate phenotypic endpoints recapitulates zebrafish developmental cascades. Toxicol Appl Pharmacol 314:109–117, PMID: 27884602, https://doi.org/10.1016/j.taap.2016.11.010.

Zhao G, Marceau R, Zhang D, Tzeng JY. 2015. Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. Genetics 199(3):695–710, PMID: 25585620, https://doi.org/10.1534/genetics.114.171686.